Solution of Exercise Sheet 4

Exercise 1   (Amazon Web Services)

1. What is the slashdot effect and how can it's negative outcome be avoided?

2. What is a region?

3. What is an availability zone?

4. What is the maximum number of instances, an EBS volume can be connected with?

5. According to which principle works the storage service S3?

   ☐ block-based storage service   ☒ object-based storage service

6. According to which principle works the service EBS?

   ☒ block-based storage service   ☐ object-based storage service

7. Which storage services require the user/customer to choose and deploy a file system?

   ☒ block-based storage service   ☐ object-based storage service

8. What is the purpose of the Access Control List (ACL) of the S3 service?

   Users can specify for all their objects and buckets the access privileges via the ACL.

9. How can users/customers increase the read and write performance (MB/s) of EBS storage?

   By using software RAID-0.

10. How can users/customers increase the availability of EBS storage?

    By using software RAID-1 or a distributed file system like Ceph, HDFS or GlusterFS.

Exercise 2   (Virtual Desktop in the Cloud)

Cloud infrastructure services like EC2 can be used to realize a virtual desktop in the Cloud.

1. Start an instance with the Linux (no matter which Linux distribution) operating system and another instance with Microsoft Windows (no matter which revision) in EC2 or in another cloud infrastructure service. Choose the infrastructure service you prefer.

2. Setup a remote desktop solution for both, the Linux instance and the Microsoft Windows instance. Different protocols and solutions exist. Examples for remote desktop protocols are Remote Desktop Protocol (RDP), Virtual Network Computing (VNC), X Window System (X11) and NX NoMachine. Choose for both operating systems the remote desktop solutions you prefer.

TBD.

Exercise 3   (Infrastructure Design Aspects)

Imagine that you want to realize a cluster of virtual server instances in EC2.

1. You can place all instances in a single region or place them in different regions. Name an advantage and a drawback of this method.

   - *A single region:*

     - Benefits: Traffic between instances is cost-free if these instances are located inside the same region. Latency is better compared with communication over region boundaries.

     - Drawback: If the region which contains your instances fails, all your instances are gone.

   - *Different regions:*

     - Benefit: Better reliability. If one region fails, not all your instances are gone.

     - Drawbacks: You need to pay the traffic between the instances when they communicate over region boundaries. Latency is worse compared with communication inside regions.

2. You can place all instances in a single availability zone or place them in different availability zones. Name an advantage and a drawback of this method.

   - *A single availability zone:*

     - Benefits: Latency is better compared with communication over availability zone boundaries.

     - Drawback: If the availability zone which contains your instances fails, all your instances are gone.

- *Different availability zones:*

  – Benefit: Better reliability. If one availability zone fails, not all your instances are gone.

  – Drawbacks: Latency is worse compared with communication inside availability zones.

Exercise 4   (Create a simple Web Server)

Implement a highly available web server with the Amazon Web Services (AWS).

Start an server instance with the Linux operating system in EC2 (Elastic Compute Cloud). Install in the instance a web server software like Apache or nginx. The web server should be accessible from the internet and show a web page. The site content does not matter.

TBD.

Exercise 5   (Persistent Storage)

Valuable data should be stored in a persistent storage. In order to satisfy this condition, create an elastic block store volume with the file system `ext3` and attach the volume to your server instance. Move the web page(s) of your web server on the volume, adjust your web server configuration file and restart the web server software.

TBD.

Exercise 6   (Realize a High Throughput Cluster)

Start a second server instance in EC2, which contains the Linux operating system, the web server software and has a persistent storage volume attached. Try to simplify the start of the second instance by using an image and snapshots.

Create an elastic load balancer which distributes incoming requests to the web server instances. Test if your virtual web server infrastructure works. If you request the load balancer multiple times, both web server instances should be addressed.

TBD.

Exercise 7   (Different persistent Storage Solutions for the High Throughput Cluster)

For exercise 6, you implemented with the infrastructure services of the Amazon Web Services (AWS), a highly available High Throughput Cluster of virtual web servers. One part of this task was to attach a persistent storage volume to each web server instance. All storage volumes contained identical data.

Not for all possible scenarios, it is the optimal approach, when all instances have EBS volumes with identical content attached.

1. Investigate other approaches and summarize their advantages and drawbacks. The focus of your investigation should be the aspects availability, cost, throughput, scalability and complexity of the different approaches.

   Several options exist to solve this problem. One option is creating a distributed file system on top of the EBS volumes. Such a solution scales well, but is it not so simple to increase the cluster with new nodes. Several distributed file systems like GlusterFS provide internal redundancy, which improves reliability a lot. Another benefit is that for such a solution the traffic between instances and a storage service would be cost-free.

   Another option is using a network protocol like the network file system (NFS). One instance acts as server and the other instances are the clients. When using such a solution, it is simple to add new clients, but the server may become a bottleneck and NFS does not provide internal redundancy. Also for this solution the traffic between the instances and the storage would be cost-free.

   Another option is using a object-oriented storage service like S3. Such a service provides a good availability level. Mounting a storage service is possible with solutions like s3fs-fuse, which is a FUSE-based file system backed by Amazon S3. A drawback of this option is that the traffic between instances and a storage service like S3 is not for cost-free.

2. Implement at least one of the solutions you investigated.

   TBD.

Exercise 8   (Financial Aspects)

As CEO of a young IT company (start-up), you face the challenge of implementing the IT infrastructure, your business requires. Since your products are primarily web services, you require a highly available and elastic infrastructure, because when you go live, you will have only few customers and it is unclear, when the number of customers will increase. The following components are absolutely required:

- 2 web servers. Each instance needs to be equipped with at least 2 (virtual or physical) CPU (cores) and 4 GB main memory.

- A load balancer to distribute incoming requests to the web servers, according to their utilization.

- 25 TB, persistent storage for the web servers.

- A relational database (dedicated, physical server or instance or cloud service) for user data and log data which stores 10 GB of data.

- 5 GB internal data transfer per month.

- 100 GB external data transfer per month.

- Every week you do a full backup of the 25 TB persistent storage. Backups which are older than 4 weeks, will be erased. This means you need to store up to 100 TB backup data.

The venture capital, you collected to start a business from your family, friends and banks is very limited. For this reason, try to implement the required infrastructure as inexpensive as possible. Your infrastructure should be able to handle in an elastic way, a growing or decreasing number of customers in the future.

Your task is to do a cost analysis for these implementation scenarios:

1. Calculate the monthly costs, when you realize the required infrastructure with the Amazon Web Services (AWS).

2. Calculate the monthly costs, when you realize the infrastructure with traditional hosting offerings. Check the websites of at least a single well-known hosting provider to investigate their pricing.

3. Calculate the monthly costs, when you buy the required physical hardware and run it for yourself in your company. Check the websites of at least a single well-known hardware vendor to investigate the pricing.

TBD.

Exercise 9   (Interacting with Storage Services)

1. Four HTTP methods are enough to work with resources inside storage services like S3 or Google Cloud Storage. Write the HTTP methods into the table.

HTTP method	Description
PUT	Create or replace resource
GET	Request resource
POST	Append something to a resource
DELETE	Erase resource

2. Why is it recommendable, that storage services do not only implement support for the four HTTP methods of subtask 1, but also for the HTTP method HEAD?

   HEAD *requests the metadata of a resource. It does request information about e.g. an object without the need to transfer the entire object.*

Exercise 10 (AWS Import/Export vs. Internet)

Your local time in Frankfurt am Main is Monday 09:00 (UTC+1). You need to copy 3 TB of data into the storage service S3. You have two options:

- **Scenario 1:** You immediately start at 09:00 (UTC+1) to upload the 3 TB of data to S3 via the internet. Consider the data rate between your computer and S3 is 100 Mbit/s.

- **Scenario 2:** You use the AWS Import/Export service [1]. Therefore you copy the data to a HDD, which is connected via USB 3.0. The transfer rate (for write) is 125 MB/s.

  After you copied the data, you pack the HDD into a parcel and send it via a package delivery company to Amazon. DHL, UPS and FedEx can deliver a parcel from Frankfurt am Main in less than 24 hours to most places in Europe.

  You need 15 Minutes to put the HDD into a parcel and another 15 Minutes to bring the parcel to the branch office of a package delivery company.

  The parcel must arrive at the branch office of the package delivery company no later than 16:30 (UTC+1) to arrive at Amazon in Ireland at 9:00 (UTC) the next working day.

  An Amazon employee needs to copy the data from the HDD to the S3 service. The transfer rate of the HDD (for read) is 150 MB/s.

  Consider 3 hours additional overhead for the in-house mail at Amazon to ship the HDD to the correct employee.

Calculate. . .

1. for the first scenario, how long it takes until the data is copied to S3.

2. for the second scenario, how long it takes until the data is copied to S3.

3. the data rate of the second scenario.

(For all subtasks, the calculation steps must be visible.)

[1] http://aws.amazon.com/importexport/

Scenario 1:

$$\frac{100 \, \text{Mbps}}{8} = 12.5 \, \text{MB/s} = 12.5 * 10^6 \, \text{Byte/s}$$

$$\frac{3 * 10^{12} \, \text{Byte}}{12.5 * 10^6 \, \text{Byte/s}} = 0.24 * 10^6 \, \text{s} = 240,000 \, \text{s}$$

$$\implies \frac{240,000 \, \text{s}}{60} = 4,000 \, \text{m} \implies \frac{4,000 \, \text{m}}{60} = 66.\overline{6} \, \text{h}$$

$\implies$ *the data transmission requires 2 days, 18 hours, 40 minutes*

Scenario 2:

Day 1, 9:00 (UTC+1)

Write data: $\frac{3,000,000 \, \text{MB}}{125 \, \text{MB/s}} = 24,000 \, \text{s} = 400 \, \text{m} = 6 \, \text{h}, 40 \, \text{m}$

Day 1, 15:40 (UTC+1)

30 minutes to put the HDD into a parcel and to bring it to the package delivery company.

Day 1, 16:10 (UTC+1)

Because the packet arrived at the package delivery company before 16:30 (UTC+1), it is delivered to Amazon the next working day at 9:00 (UTC).

Day 2, 9:00 (UTC) = 10:00 (UTC+1)

3 hours are required to ship the HDD via in-house mail at Amazon to the correct employee.

Day 2, 12:00 (UTC) = 13:00 (UTC+1)

Read data: $\frac{3,000,000 \, \text{MB}}{150 \, \text{MB/s}} = 20,000 \, \text{s} = 333.33\overline{3} \, \text{m} = 5 \, \text{h}, 34 \, \text{m}$

Day 2, 17:34 (UTC) = 18:34 (UTC+1)

$\implies$ *the data transmission requires 1 day, 9 hours, 34 minutes = 2,014 minutes = 120,840 seconds*

$$\frac{3 * 10^{12}\,\text{Byte}}{120,840\,\text{s}} = \frac{3,000,000 * 10^6\,\text{Byte}}{120,840\,\text{s}} = 24.8262164846 * 10^6\,\text{Byte/s}$$

$$24.8262164846 * 10^6\,\text{Byte/s} * 8 = 198,609,731.877 * 10^6\,\text{Bit/s} = approx.\ 199\,\text{Mbps}$$

The data rate of the second scenario is approx. two times better compared with the first scenario!